Generative AI-Powered Resume Analyzer: Documentation

1. Overview

This project automates resume screening using Generative AI techniques and NLP methods. It parses PDF resumes, extracts relevant fields (Name, Contact, Education, Skills, etc.), and scores each candidate on Generative AI and AI/ML experience. An Excel file is produced with the consolidated data.

2. Approach

2.1 Resume Text Extraction

- **PyPDF2 or pdfplumber**: Extract text from PDF resumes.
- **Tesseract OCR**: Fallback for scanned/image-based resumes, ensuring we handle different file formats robustly.

2.2 Generative AI Integration

- **OpenAI's GPT (ChatCompletion API)**:
    1. **Context-aware parsing**: We prompt GPT to return JSON with fields like name, phone, emails, multiple education entries, and skill sets.
    2. **Scoring function**: GPT assigns numerical scores for Generative AI Experience and AI/ML Experience. We also demonstrate a Job Match Score by comparing the resume text to a job description.

2.3 Data Cleaning & Enhancement

- **Phone Normalization**: Convert raw phone strings into consistent formats (+91-xxxxxxxxxx, (xxx) xxx-xxxx).
- **CGPA vs. Percentage**: Parse CGPA vs. % and label them accordingly.
- **Bullet-Point Formatting**: Present multi-section data (e.g., Projects, Certifications, Additional Insights) in an easy-to-read enumerated style.

2.4 Batch Processing & Caching

- **Batch/Queue Approach**: We can handle multiple resumes (up to 100 or more) by queueing filenames. This reduces the risk of hitting API limits all at once and makes the code easier to scale.
- **Caching**: For repeated runs on the same resumes, we store processed results keyed by text hash to avoid redundant API calls.

2.5 Output

- **Excel File**: Final structured data includes:
    1. Mandatory fields like Name, Contact, Education, etc.
    2. AI-driven scores (Generative AI Score, AI/ML Score).
    3. Additional sections (Supporting Information, e.g., awards, volunteering, job match score).

3. Innovative Features

- **Context-Aware GPT Extraction**: Minimizes reliance on rigid regex, adapting to varied resume formats.
- **Advanced Scoring**: Provides numerical evaluations for Generative AI, AI/ML, and an optional "Job Match Score."
- **Neat Presentation**: Bullet-point formatting ensures a readable output in Excel.

4. Future Enhancements

- **Regex Fallback**: For phone/email if GPT occasionally misses data.
- **libphonenumber**: For more robust phone formatting across regions.
- **Deeper Analytics**: Extended heuristics to interpret career potential, skill synergy, etc.

5. How to Run

Refer to the **Instructions for Running the Code** file for environment setup and execution details.

6. Conclusion

By leveraging GPT for both data extraction and scoring, this project demonstrates how Generative AI can streamline resume analysis at scale, offering a flexible, robust, and user-friendly approach for HR and recruitment processes.